

**PERBANDINGAN ALGORITMA DECISION TREE (C4.5) DAN NAÏVE BAYES PADA  
DATA MINING UNTUK IDENTIFIKASI TUMBUH KEMBANG ANAK BALITA  
(STUDI KASUS PUSKESMAS KARTASURA)**

**NASKAH PUBLIKASI**

**PROGRAM STUDI INFORMATIKA**

**FAKULTAS KOMUNIKASI DAN INFORMATIKA**



**Diajukan oleh :**

**Mila Listiana**

**Drs.Sudjalwo,M.Kom.**

**Dedi Gunawan,S.T.,M.Sc.**

**PROGRAM STUDI INFORMATIKA**

**FAKULTAS KOMUNIKASI DAN INFORMATIKA**

**UNIVERSITAS MUHAMMADIYAH SURAKARTA**

**JULI, 2015**

## HALAMAN PENGESAHAN

Publikasi ilmiah dengan judul :

**PERBANDINGAN ALGORITMA DECISION TREE (C4.5) DAN NAÏVE BAYES PADA  
DATA MINING UNTUK IDENTIFIKASI TUMBUH KEMBANG ANAK BALITA  
(STUDI KASUS PUSKESMAS KARTASURA)**

Yang dipersiapkan dan disusun oleh :

Mila Listiana

L200110047

Telah disetujui pada :

Hari : Senin

Tanggal : 13 Juli 2015

Pembimbing I

Drs. Sudjalwo, M.Kom.  
NIK : 404

Pembimbing II

Dedi Gunawan, S.T., M.S.c.  
NIK : 1303

Publikasi ilmiah ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar sarjana  
Tanggal 3 Agustus 2015

Mengetahui,

Ketua Program Studi  
Informatika



Dr. Heru Supriyono, M.Sc.

NIK : 970



UNIVERSITAS MUHAMMADIYAH SURAKARTA  
FAKULTAS KOMUNIKASI DAN INFORMATIKA  
PROGRAM STUDI INFORMATIKA

Jl. A Yani Tromol Pos 1 Pabelan Kartasura Telp. (0271)717417, 719483 Fax (0271) 714448  
Surakarta 57102 Indonesia. Web: <http://informatika.ums.ac.id>. Email: [informatika@fki.ums.ac.id](mailto:informatika@fki.ums.ac.id)

**SURAT KETERANGAN LULUS PLAGIASI**

**/A.3-II.3/INF-FKI/VII/2015**

Assalamu'alaikum Wr. Wb

Biro Skripsi Program Studi Informatika menerangkan bahwa :

Nama : MILA LISTIANA  
NIM : L200110047  
Judul : PERBANDINGAN ALGORITMA DECISION TREE (C4.5) DAN  
NAIVE BAYES PADA DATA MINING UNTUK IDENTIFIKASI  
TUMBAH KEMBANG ANAK BALITA (STUDI KASUS  
PUSKESMAS KARTASURA)  
Program Studi : Informatika  
Status : **Lulus**

Adalah benar-benar sudah lulus pengecekan plagiasi dari Naskah Publikasi Skripsi, dengan menggunakan aplikasi Turnitin.

Demikian surat keterangan ini dibuat agar dipergunakan sebagaimana mestinya.

Wassalamu'alaikum Wr. Wb

*Surakarta, 29 Juli 2015*

Biro Skripsi  
Informatika

**Adjie Sapoetra, S.Kom**



## Turnitin Originality Report

PERBANDINGAN ALGORITMA DECISION  
TREE (C4.5) DAN NAIVE BAYES PADA  
DATA MINING UNTUK IDENTIFIKASI  
TUMBUH KEMBANG ANAK BALITA  
(STUDI KASUS PUSKESMAS  
KARTASURA) by Mila Listiana

From publikasi september 2015 (publikasi)

Similarity Index	Similarity by Source
29%	Internet Sources: 18% Publications: 0% Student Papers: 23%

Processed on 28-Jul-2015 11:05 WIB

ID: 558129288

Word Count: 2871

## sources:

1 3% match (student papers from 14-Jul-2015)  
Class: publikasi  
Assignment:

Paper ID: 555649112

2 2% match (student papers from 17-Jun-2014)  
Class: publikasi maret 2014  
Assignment:  
Paper ID: 435295687

3 2% match (Internet from 14-Jun-2015)  
<http://novena90.blogspot.com/>

4 2% match (student papers from 19-Jun-2014)  
Class: publikasi maret 2014  
Assignment:  
Paper ID: 435667253

5 1% match (Internet from 16-Jun-2015)  
<http://mainaardi.blogspot.com/>

6 1% match (student papers from 08-Jul-2014)  
Class: publikasi maret 2014  
Assignment:  
Paper ID: 438403484

7 1% match (student papers from 17-Jun-2015)  
[Submitted to Universitas Dian Nuswantoro on 2015-06-17](#)

8 1% match (student papers from 10-Mar-2015)  
Class: publikasi  
Assignment:  
Paper ID: 514651061

9 1% match (Internet from 28-Feb-2015)  
<http://www.fisioterapi.web.id/2011/10/stimulasi-tumbuh-kembang-anak-untuk.html>

10 1% match (student papers from 13-Jun-2014)  
Class: publikasi maret 2014  
Assignment:  
Paper ID: 434536412

11 1% match (student papers from 11-Mar-2014)  
Class: publikasi maret 2014  
Assignment:  
Paper ID: 404749786

12 1% match (Internet from 13-Jul-2015)  
[http://eprints.dinus.ac.id/13108/1/jurnal\\_13496.pdf](http://eprints.dinus.ac.id/13108/1/jurnal_13496.pdf)

13 1% match (student papers from 10-Mar-2015)  
Class: publikasi  
Assignment:  
Paper ID: 514619482

14 1% match (Internet from 13-Jul-2015)  
<http://research.pps.dinus.ac.id/lib/jurnal/Vol%2009.2%20049-053.pdf>

15 1% match (Internet from 02-Feb-2015)  
<http://papers.gunadarma.ac.id/index.php/industry/article/download/816/797>

# **PERBANDINGAN ALGORITMA DECISION TREE (C4.5) DAN NAÏVE BAYES PADA DATA MINING UNTUK IDENTIFIKASI TUMBUH KEMBANG ANAK BALITA (STUDI KASUS PUSKESMAS KARTASURA)**

Mila Listiana, Sudjalwo, Dedi Gunawan

Informatika, Fakultas Komunikasi dan Informatika,  
Universitas Muhammadiyah Surakarta  
Email: [milalistiana7@gmail.com](mailto:milalistiana7@gmail.com)

## **ABSTRAKSI**

Puskesmas merupakan salah satu instansi kesehatan yang berada ditingkat kecamatan. Sebagai upaya instansi kesehatan dalam mendukung pelayanan yang baik diperlukan tata kerja yang tertib, rapi dan teliti sehingga akan menghasilkan informasi yang cepat, tepat dan akurat. Dalam instansi kesehatan banyak data yang dari setiap tahunnya bertambah. Salah satunya yaitu data tumbuh kembang balita. Akan tetapi dengan sekian banyaknya data tumbuh kembang balita semakin sulit juga data tersebut dipelajari lebih lanjut, dan umumnya data tersebut hanya digunakan sebagai arsip saja. Pemanfaatan teknik data *mining* diharapkan dapat membantu dalam mengatasi tumbuh kembang balita yang kurang baik saat ini. Pada penelitian ini kami membandingkan teknik klasifikasi dari kinerja metode *decision tree(C4.5)* dan *naive bayes*. Atribut yang digunakan terdiri dari *Gender*, Usia, Berat badan, Waktu, Wilayah dan Tumbuh kembang. Dengan menggunakan masing-masing data *training* dan data *testing* sebanyak 304 data. Hasil dari penelitian yang dilakukan, berdasarkan dari nilai *accuracy* maupun *recallnya* *naive bayes* lebih tinggi dibandingkan dengan *decision tree* yaitu dengan nilai *accuracy* 75,66% untuk *decision tree* dan 76,97% untuk *naive bayes*. Untuk nilai *recall*-nya *naive bayes* lebih unggul yaitu 96,89% dibandingkan *decision tree* 89,78%. Meskipun dalam penelitian ini tingkat *Precision*-nya lebih tinggi *decision tree* yaitu 85,23% dibandingkan *naive bayes* 84,17%. Hasil akhir dari penelitian ini adalah metode *naive bayes* lebih baik digunakan dari pada metode *decision tree* dengan nilai total 250,67% untuk *decision tree* dan 258,03% untuk *naive bayes*.

**Kata kunci :** *Data Mining, Perbandingan Algoritma, Decision Tree(c4.5), Naïve Bayes.*



# **COMPARISON OF DECISION TREE ALGORITHM (C4.5) AND NAIVE BAYES ON DATA MINING FOR IDENTIFICATION GROWTH AND DEVELOPMENT CHILD TODDLER (CASE STUDY PUSKESMAS KARTASURA)**

Mila Listiana, Sudjalwo, Dedi Gunawan

Informatics, The Faculty Of Communication And Informatic,  
Muhammadiyah University Of Surakarta  
Email:[milalistiana7@gmail.com](mailto:milalistiana7@gmail.com)

## **ABSTRACT**

Puskesmas is one of health agencies located in the sub-district. Health agencies as an effort in support of good service necessary management of working order , neat and thoroughly so that will produce rapid information , proper and accurate. In health agency lots of data from which every year increase. One of them is growth and data development of toddler. But with so much data growth and development of toddler also getting harder these data studied further, and generally the data is only used as an archive only. Utilization of data mining techniques are expected to help in the growth and development of toddler overcome unfavorable currently. At this research we compared the techniques classification of performance decision tree method (C4.5) and Naive Bayes. Attributes are used consisted of Gender, Age, Weight, Time, Regional and growth and development. By using each training data and testing data as much as 304 Data. Results of research conducted, based on the value of accuracy and the recall of Naive Bayes is higher than the decision tree is to the value accuracy of 75.66% for decision tree and 76.97% for Naive Bayes. For the value recall of Naive Bayes its more superior is 96.89% compared decision tree is 89.78%. Although in this research the level Precision was higher decision tree is 85.23% compared Naive Bayes is 84.17%. The final result of this research is Naive Bayes method is better used than in the decision tree method with a total value of 250.67% for decision tree and 258.03% for Naive Bayes.

***Keywords : Data Mining, Comparison Algorithm, Decision Tree (C4.5), Naïve Bayes.***

## 1. PENDAHULUAN

Data *mining* merupakan serangkaian proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. Data *mining* juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Dalam data *mining* terdapat banyak teknik dalam pengerjaannya, diantaranya yaitu algoritma *naive bayes*, *decision tree*, jaringan saraf tiruan dan masih banyak lainnya.

Puskesmas merupakan salah satu instansi kesehatan yang berada ditingkat kecamatan. Sebagai upaya instansi kesehatan dalam mendukung pelayanan yang baik diperlukan tata kerja yang tertib, rapi dan teliti sehingga akan menghasilkan informasi yang cepat, akurat dan tepat waktu sesuai kebutuhan. Sedangkan dalam instansi kesehatan tentunya banyak data yang dari setiap tahunnya bertambah, sehingga dari data yang banyak dan menumpuk tersebut tidak dapat dipelajari lebih lanjut dan data tersebut hanya digunakan sebagai arsip saja. Salah satu data yang perlu diperhatikan yaitu mengenai tumbuh kembang balita, yang dalam penelitian saat ini banyak tumbuh kembang balita yang kurang baik. Untuk memudahkan pihak instansi dalam mengolah data yang banyak tersebut, maka dibutuhkan sistem yang dapat menghasilkan sebuah keputusan untuk

tumbuh kembang balita. Salah satunya yaitu menggunakan teknik data *mining*. Sehingga data yang banyak tersebut dapat dimanfaatkan secara optimal.

Dengan dimanfaatkannya data tumbuh kembang balita diharapkan dapat membantu mengatasi tumbuh kembang balita yang kurang baik saat ini. Dan dengan metode *mining* diharapkan dapat membantu pihak instansi kesehatan dalam memantau kesehatan tumbuh kembang balita dari setiap periodenya dengan lebih mudah dan cepat. Selain itu juga dapat digunakan sebagai sarana pengambilan keputusan untuk lebih meningkatkan pelayanan kesehatan yang tumbuh kembang balita kurang baik.

Dari permasalahan diatas dapat diambil alternatif dengan cara memanfaatkan teknik data *mining* dengan membandingkan 2 metode untuk identifikasi tumbuh kembang balita menggunakan metode *decision tree* (c4.5) dan *naive bayes*. Dengan harapan setelah diolah dengan data *mining* dapat membantu menemukan informasi dalam identifikasi tumbuh kembang balita.

## 2. TINJAUAN PUSTAKA

### 2.1 Telaah Penelitian

Pada penelitian Maulina (2014) mengemukakan bahwa dengan adanya perubahan tertentu pada DNA sel payudara dapat mengakibatkan kanker payudara.

Untuk mengetahui kanker payudara dapat dilakukan dengan cara mendeteksi sel tersebut dengan uji laboratorium. Dalam penganalisaan digunakan 9 atribut acuan. Salah satunya yaitu menggunakan teknik data *mining* untuk menganalisis pola atau pengetahuan dari sekumpulan data secara otomatis. Pada sistem ini, teknik klasifikasi menggunakan metode *decision tree* algoritma (*ID-3*). Dimana algoritma tersebut merupakan algoritma yang paling dasar dalam *decision tree* dan merupakan metode untuk membangun *decision tree* dalam mencari solusi. Dari hasil uji coba yang dilakukan sistem dapat memberikan keputusan yang cukup baik dalam klasifikasi kanker payudara yang terbagi menjadi 2 kelas yaitu jinak dan ganas dengan hasil akurasi mencapai 98.5%.

Sedangkan dalam penelitian Munawaroh (2013) menyatakan pemilihan penjurusan di SMA Negeri 2 Bangkalan merupakan masalah yang selalu dihadapi oleh guru. Terutama masalah keterlambatan nilai siswa dari wali kelas ditambah lagi banyaknya masalah siswa kelas x sehingga proses penjuruan kurang tepat. Dari permasalahan diatas diperlukan sistem terkomputerisasi yang dapat mengatasi masalah tersebut. Oleh karena itu dibangun aplikasi untuk mengidentifikasi penjurusan SMA. Aplikasi ini menggunakan perbandingan algoritma ID3 dan C5.0. Namun belum

diketahui algoritma mana yang lebih unggul kinerjanya. Oleh karena itu kedua algoritma tersebut perlu dibandingkan. Dalam penelitian ini menggunakan 150 data *training* dan 50 data *testing*. Hasil penelitian didapat nilai *precision* 94,87%, nilai *recall* 97,37% dan nilai *accuracy* 94%, selain itu nilai *error rate* 6% untuk algoritma C5.0. Hasil akhir adalah algoritma C5.0 lebih unggul dari pada ID3.

## **2.2 Landasan Teori**

### **a. Tumbuh Kembang Balita**

Tumbuh kembang sebenarnya mencakup dua peristiwa yang sifatnya berbeda tetapi saling berkaitan, yaitu pertumbuhan dan perkembangan. Pertumbuhan merupakan perubahan dalam jumlah besar, ukuran yang bisa diukur dengan ukuran berat, panjang. Sedangkan perkembangan merupakan bertambahnya kemampuan dalam struktur dan fungsi tubuh yang lebih kompleks, dari hasil proses pematangan (Kania, 2006)

### **b. Data Mining**

Data *mining* adalah serangkaian proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar (Prasetyo, 2012)

### **c. Decision Tree**

*Decision tree* merupakan metode yang paling efisien untuk menyaring sesuatu lewat pohon keputusan apakah suatu data lolos atau tidak terhadap



saringan dengan proses yang cukup cepat. Ada beberapa tahap dalam membuat sebuah pohon keputusan yaitu: (Saputra, 2014)

1. Menyiapkan data *training* yang sudah dikelompokkan ke dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon. Akar akan diambil dari atribut yang terpilih, yaitu dengan cara menghitung nilai *gain* dari masing-masing atribut. Nilai *gain* yang tertinggi akan menjadi akar pertama. Sebelum menghitung nilai *gain* dari atribut, hitung nilai *entropy*.

Hal yang harus dilakukan dalam metode *decision tree* adalah menghitung *entropy* dan *information gain*. Untuk menghitung nilai *entropy* digunakan rumus:

$$Entropi(y) = -p_1 \log_2 p_1 - p_2 \log_2 p_2 - \dots - p_n \log_2 p_n \dots$$

Dimana  $p_1, p_2, \dots, p_n$  masing-masing menyatakan proporsi kelas 1, kelas 2, ..., kelas n dalam output.

Sedangkan untuk menghitung nilai *information gain* menggunakan rumus:

$$gain(y, A) = entropi(y) - \sum_{c \in \text{nilai}(A)} \frac{y_c}{y} entropi(y_c)$$

Dimana atribut *information gain* tertinggi akan dijadikan *node*.

#### d. Algoritma C4.5

*Algoritma C4.5* merupakan algoritma pengembangan dari algoritma ID3 yang digunakan untuk membentuk pohon keputusan. Pohon keputusan dapat membagi kumpulan data yang besar menjadi himpunan-himpunan *record* yang lebih kecil dengan menerapkan serangkaian aturan keputusan (Sofiana, 2012)

Secara umum *algoritma C4.5* untuk membangun pohon keputusan adalah sebagai berikut:

1. Pilih atribut sebagai akar
2. Buat cabang untuk masing-masing nilai
3. Bagi kasus dalam cabang
4. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama (Abidin, 2011)

#### e. Naive Bayes

*Naive bayes* merupakan teknik prediksi berbasis probabilitas sederhana yang berdasar pada penerapan aturan bayes dengan asumsi ketidaktergantungan yang kuat. Selain itu *naive bayes* juga dapat menganalisa variabel-variabel yang paling mempengaruhinya dalam bentuk peluang (Prasetyo, 2012).

Berikut merupakan persamaan dari teorema *Bayes*.

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \dots$$

Keterangan:

X : data dengan class yang belum diketahui

H : hipotesis data X merupakan suatu class spesifik

$P(H|X)$  : probabilitas akhir bersyarat suatu hipotesis H terjadi jika diberikan bukti X terjadi

$P(H)$  : Probabilitas awal (*priori*) hipotesis H terjadi tanpa memandang bukti apapun

$P(X|H)$  : probabilitas sebuah bukti X terjadi akan mempengaruhi hipotesis H.

$P(X)$  : probabilitas awal (*priori*) bukti X terjadi tanpa memandang hipotesis/ bukti yang lain.

#### f. PHP

PHP atau *Hypertext Preprocessor* adalah bahasa skrip yang dapat ditanamkan atau disisipkan kedalam HTML. PHP banyak dipakai untuk memrogram situs web dinamis.

#### g. MySQL

MySQL adalah sebuah implementasi dari sistem manajemen basisdata relasional yang didistribusikan secara gratis dibawah lisensi GPL (*General Public License*). Setiap pengguna dapat secara bebas menggunakan MySQL, namun dengan batasan perangkat lunak tersebut tidak

boleh dijadikan produk turunan yang bersifat komersial.

#### h. Precision, Recall, Accuracy

Nilai *precision* merupakan proporsi jumlah kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya. Nilai *recall* merupakan proporsi jumlah kasus positif yang sebenarnya yang diprediksi positif secara benar pula. Sedangkan nilai *accuracy* merupakan prosentase jumlah data yang diklasifikasi secara benar secara keseluruhan (Andriani, 2013)

### 3. METODE PENELITIAN

#### 3.1 Pengumpulan Data

Berdasarkan masalah dan kebutuhan data *mining* yang ada maka data-data yang dibutuhkan adalah:

Tabel 3.1daftar atribut keseluruhan

Atribut	Nilai Atribut	Tipe
Tumbuh Kembang	Normal, gizi kurang, gizi buruk, gizi lebih	polynomial
<i>Gender</i>	Laki-laki, perempuan	Binomial
Usia	3, 7, 8, 9, 10 dsb	Real
Berat Baddan	5; 6; 6,9 dsb	Real
Waktu	Januari, februari, maret dsb	Polynomial
Wilayah	Bakalan, Beteng sari, Blaterasi, dsb	Polynomial

### 3.2 Penentuan Atribut

Berdasarkan masalah dan kebutuhan tahap yang pertama adalah menentukan atribut dengan menyeleksi data keseluruhan. Ditetapkan atribut yang digunakan sebagai variabel dependen dan independen.

Tabel 3.2 penentuan variabel dependen dan independen

Atribut	Variabel
Tumbuh Kembang	Y
Gender	X1
Usia	X2
Berat Baddan	X3
Waktu	X4
Wilayah	X5

Dalam penelitian ini atribut tumbuh kembang dijadikan sebagai variabel dependen (Y) sedangkan atribut *gender*, usia, berat badan, waktu dan wilayah sebagai variabel independen (X1), (X2), (X3), (X4), dan (X5).

### 3.3 Pengelompokan Data

Setelah selesai menentukan atribut yang akan digunakan kemudian nilai dari atribut tersebut dikelompokkan untuk pengklasifikasian data *mining*.

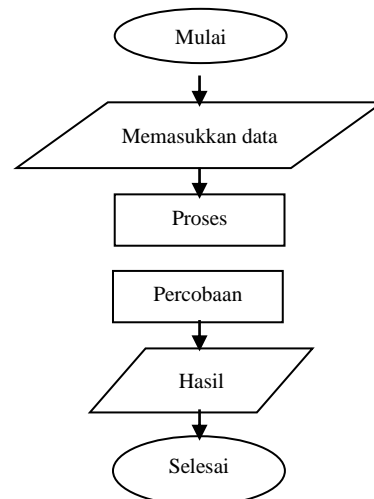
Tabel 3.3 pengelompokan data

Nilai atribut	Kelas
Normal, gizi kurang, gizi buruk, gizi lebih	Normal, Gizi kurang, Gizi buruk, Gizi lebih
Laki-laki, perempuan	Laki-laki, Perempuan

3, 7, 8, 9, 10 dsb	usia $\leq$ 15, 15<usia $\leq$ 30, 30<usia $\leq$ 45, usia>45
5; 6; 6,9 dsb	berat $\leq$ 10, 10<berat $\leq$ 20, berat>20
Januari, february, maret dsb	Kuartal I, Kuartal II, Kuartal III, Kuartal IV
Bakalan, Beteng sari, Blaterasi, dsb	Wilayah I, Wilayah II, Wilayah III, Wilayah IV, Wilayah V, Wilayah VI

### 3.4 Perancangan Sistem

Dalam sistem yang akan dibangun terdapat beberapa proses utama, antara lain:



Gambar 1. Rancangan Sistem

## 4. HASIL DAN PEMBAHASAN

### 4.1 Implementasi Dengan Perhitungan *Decision Tree*.

#### 4.1.1 Menentukan *Root Node*

*Root Node* merupakan node paling atas. Untuk menentukan *root* atau akar pada *decision tree* adalah dengan menentukan nilai *information gain* dari setiap data yang telah ditentukan

berdasarkan atribut yang telah ditentukan sebelumnya.

Atribut yang dipilih sebagai akar adalah atribut yang memiliki nilai *information gain* yang paling tinggi.

Dari hasil perhitungan didapatkan nilai *information gain* tertinggi seperti tabel 4.1.

**Tabel 4.1** Nilai Tumbuh Kembang

Nilai <i>Information Gain</i>	
Gender	0,004
Usia	0,138
Berat Badan	0,284
Waktu	0,015
Wilayah	0,034

Berdasarkan tabel 4.1 didapatkan nilai *information gain* tertinggi yaitu atribut berat badan dengan nilai 0,284 oleh karena itu variabel berat badan dijadikan sebagai *root*(akar).

#### 4.1.2 Menentukan internal node

##### 4.1.2.1 Menentukan internal node pada berat badan $10 < \text{berat} \leq 20$

Dalam mencari internal node pada berat badan  $10 < \text{berat} \leq 20$  didapatkan hasil perhitungan dalam mencari nilai *information gain* dari setiap atribut seperti tabel 4.2.

**Tabel 4.2.** Nilai *Information Gain*

Nilai <i>Information Gain</i>	
Gender	0,004
Usia	0,047
Waktu	0,033

Wilayah	0,027
---------	-------

Dari hasil tabel 4.2 diperoleh hasil *information gain* tertinggi adalah usia dengan nilai 0,047 oleh karena itu atribut *usia* dijadikan perantara pada berat  $10 < \text{berat} \leq 20$

##### 4.1.2.2 Menentukan internal node pada berat badan berat > 20

Dalam mencari internal node pada berat badan berat > 20 didapatkan hasil *information gain* seperti pada tabel 4.3.

**Tabel 4.3.** Nilai *Information Gain*

Nilai <i>Information Gain</i>	
Gender	0,467
Usia	0,000
Waktu	0,016
Wilayah	0,360

berdasarkan tabel 4.3 didapatkan nilai *information gain* tertinggi adalah *gender* dengan nilai 0,467 oleh karena itu atribut *gender* dijadikan perantara pada berat > 20

##### 4.1.2.3 Menentukan internal node pada berat badan berat ≤ 10

Dalam mencari internal node pada berat badan berat ≤ 10 didapatkan hasil nilai *information gain* seperti pada tabel 4.4.

**Tabel 4.4.** Nilai *Information Gain*

Nilai <i>Information Gain</i>	
Gender	0,005
Usia	0,107

Waktu	0,273
Wilayah	0,151

Berdasarkan tabel 4.4 diperoleh hasil information *gain* tertinggi adalah waktu dengan nilai 0,273. Oleh karena itu variabel waktu digunakan sebagai penghubung pada berat $\leq$ 10.

## 4.2 Implementasi Dengan Perhitungan *Naive bayes*.

### 4.2.1 Perhitungan Dengan Data Training

Dalam menentukan tumbuh kembang balita berdasarkan data *training* untuk sampel “Laki-laki, usia  $30 < \text{usia} \leq 45$ , berat  $10 < \text{berat} \leq 20$ , waktu Kuartal II, wilayah berasal dari Wilayah I” didapatkan hasil seperti pada tabel 4.5

Tabel 4.5.Perhitungan HMAP

HMAP	
Normal	0,011257869
Gizi Kurang	0,000324159
Gizi Buruk	0,000000000
Gizi Lebih	0,000162419

Berdasarkan tabel 4.5 nilai tertinggi merupakan hasil dari keputusan tumbuh kembang yaitu Normal dengan nilai 0,011257869.

### 4.2.2 Perhitungan Dengan Data Testing

Dalam menentukan tumbuh kembang balita berdasarkan data *testing* untuk sampel “Perempuan, usia  $15 < \text{usia} \leq 30$ , berat $\leq$ 10, waktu Kuartal I, wilayah berasal dari Wilayah IV” didapatkan hasil seperti pada tabel 4.6.

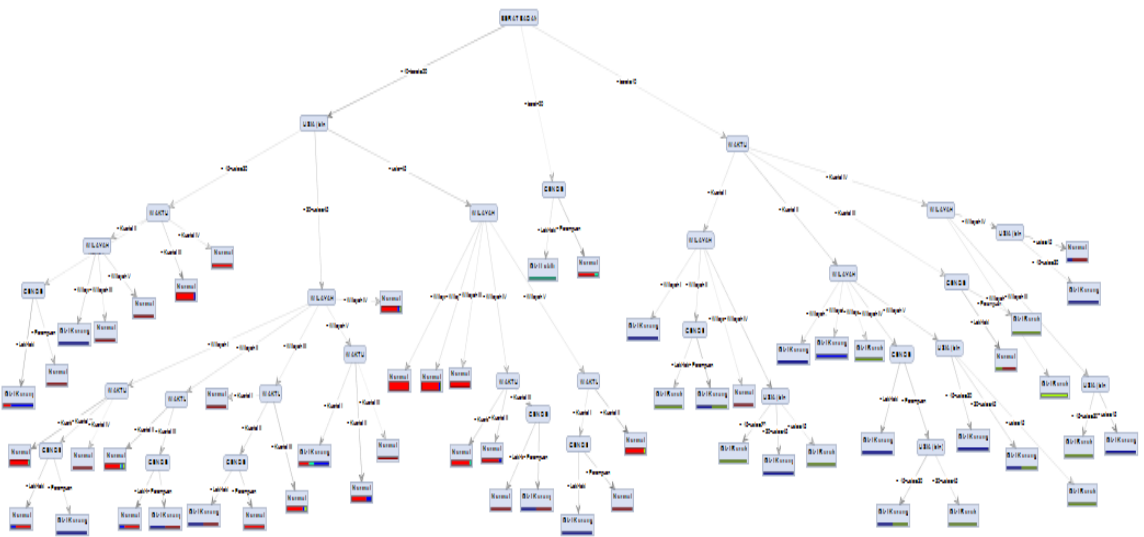
Tabel 4.6.Perhitungan HMAP

HMAP	
Normal	0,000191269
Gizi Kurang	0,000595988
Gizi Buruk	0,001728000
Gizi Lebih	0,000000000

Berdasarkan tabel 4.6 nilai tertinggi merupakan hasil dari keputusan tumbuh kembang yaitu Gizi Buruk dengan nilai 0,001728000.

## 4.3 Implementasi Decision Tree menggunakan Rapid Miner 6

Rancangan proses menggunakan *decision tree* dengan *rapid miner 6* menghasilkan sebuah skema pohon keputusan untuk mengetahui karakteristik tumbuh kembang anak balita berdasarkan atribut yang sudah ditentukan. Hasil pohon keputusan seperti pada gambar 4.1



Berdasarkan hasil pohon keputusan dapat diketahui bahwa berat badan memiliki pengaruh paling tinggi dalam menentukan tumbuh kembang balita. Hal ini ditunjukkan dengan atribut berat badan sebagai *root node*. Kemudian untuk *gender* dan waktu sebagai internal *node* pertama.

#### 4.4 implementasi Naive Bayes Menggunakan Rapid Miner 6

Rancangan proses menggunakan *naive bayes* dengan rapid miner 6 menghasilkan sebuah prediksi dalam menentukan tumbuh kembang balita berdasarkan atribut yang sudah ditentukan sebelumnya. Hasil prediksi dapat dilihat pada gambar 4.2

Row No.	confidence <sub>1</sub>	confidence <sub>2</sub>	confidence <sub>3</sub>	confidence <sub>4</sub>	prediction <sub>1</sub>	GENDER	USA (in)	BERAT (kg)	WAKTU	WILAYAH
1	0.023	0.008	0.005	0.873	Normal	Laki-laki	usa-45	10-berats-2	Kuatal	Wilayah V
2	0.188	0.082	0.014	0.716	Normal	Laki-laki	usa-45	10-berats-2	Kuatal	Wilayah V
3	0.803	0.088	0.006	0.873	Normal	Laki-laki	usa-45	10-berats-2	Kuatal	Wilayah V
4	0.214	0.034	0.018	0.734	Normal	Perempuan	usa-45	10-berats-2	Kuatal	Wilayah V
5	0.313	0.000	0.003	0.854	Normal	Laki-laki	usa-45	10-berats-2	Kuatal	Wilayah V
6	0.336	0.000	0.653	0.011	Go Buruk	Laki-laki	usa-15	berats-10	Kuatal	Wilayah IV
7	0.503	0.000	0.450	0.047	Go Kurang	Laki-laki	usa-45	berats-10	Kuatal	Wilayah IV
8	0.605	0.000	0.282	0.113	Go Kurang	Perempuan	usa-45	berats-10	Kuatal	Wilayah II
9	0.152	0.030	0.028	0.790	Normal	Laki-laki	usa-45	10-berats-2	Kuatal	Wilayah IV
10	0.582	0.000	0.385	0.023	Go Kurang	Perempuan	usa-15	berats-10	Kuatal	Wilayah IV
11	0.214	0.034	0.018	0.734	Normal	Perempuan	usa-45	10-berats-2	Kuatal	Wilayah V
12	0.477	0.000	0.483	0.040	Go Buruk	Perempuan	usa-45	berats-10	Kuatal	Wilayah IV
13	0.534	0.000	0.383	0.082	Go Kurang	Perempuan	usa-45	berats-10	Kuatal	Wilayah IV
14	0.028	0.032	0.011	0.932	Normal	Laki-laki	usa-45	10-berats-2	Kuatal	Wilayah IV
15	0.026	0.032	0.011	0.932	Normal	Laki-laki	usa-45	10-berats-2	Kuatal	Wilayah V
16	0.214	0.034	0.018	0.734	Normal	Perempuan	usa-45	10-berats-2	Kuatal	Wilayah V
17	0.080	0.000	0.013	0.906	Normal	Perempuan	usa-45	10-berats-2	Kuatal	Wilayah V
18	0.013	0.000	0.005	0.983	Normal	Perempuan	usa-45	10-berats-2	Kuatal	Wilayah III
19	0.024	0.028	0.007	0.941	Normal	Perempuan	usa-45	10-berats-2	Kuatal	Wilayah I
20	0.018	0.054	0.002	0.926	Normal	Laki-laki	usa-45	10-berats-2	Kuatal	Wilayah I
21	0.021	0.023	0.003	0.953	Normal	Perempuan	usa-45	10-berats-2	Kuatal	Wilayah I
22	0.026	0.032	0.011	0.932	Normal	Laki-laki	usa-45	10-berats-2	Kuatal	Wilayah IV

Gambar 4.2 Hasil *Naive Bayes* Untuk  
Prediksi Tumbuh Kembang Balita

Berdasarkan gambar 4.2 dapat diketahui bahwa prediksi tumbuh kembang balita menggunakan data uji terdapat pada kolom ke-6 dengan memberikan hasil keputusan tumbuh kembang balita.

## 4.5 Analisis Sistem

Berdasarkan pada perhitungan yang sudah didapat oleh karena itu menghasilkan

sebuah sistem tumbuh kembang balita dengan membandingkan hasil kinerja dari metode *decision tree* dan *naive bayes* yang telah diuji menggunakan browser *google chrome* dan mendapat pengujian sebagai berikut:

#### 4.5.1 Halaman Home

Pada halaman home ini merupakan halaman utama yang terdapat 2 menu utama yaitu metode *decision tree* dan *naive bayes*, serta terdapat form yang dapat dimasukkan nilai secara manual oleh user dan terdapat button proses untuk menampilkan hasil prediksi. Tampilan home dapat dilihat pada gambar 4.3



Gambar 4.3 Tampilan Home

Pada gambar 4.4 menunjukkan hasil proses prediksi pada sistem setelah dijalankan.



Gambar 4.4 Tampilan Proses



## 5. KESIMPULAN

Berdasarkan hasil penelitian dari perbandingan 2 metode dalam identifikasi tumbuh kembang balita tersebut, dapat diambil kesimpulan bahwa:

- 5.1 Berdasarkan hasil perbandingan metode *naive bayes* lebih baik digunakan dari pada metode *decision tree* dengan nilai total 250,67% untuk *decision tree* dan 258.03% untuk *naive bayes*.
- 5.2 Berdasarkan metode *decision tree* atribut berat badan merupakan atribut yang paling berpengaruh dalam penelitian ini. Dibuktikan dengan atribut berat badan menduduki *root node*.
- 5.3 Berdasarkan dari nilai *accuracy* maupun *recallnya naive bayes* lebih tinggi dibandingkan dengan *decision tree* yaitu dengan nilai *accuracy* 75,66% untuk *decision tree* dan 76,97% untuk *naive bayes*. Untuk nilai *recall-nya naive bayes* lebih unggul yaitu 96,89% dibandingkan *decision tree* 89,78%. Meskipun dalam penelitian ini tingkat *Precision-nya* lebih tinggi *decision tree* yaitu 85,23% dibandingkan *naive bayes* 84,17%.

## Daftar Pustaka

- Abidin, A. Z. Z. (2011). "*Implementasi Algoritma C4. 5 untuk Menentukan Tingkat Bahaya Tsunami.*" Seminar Nasional : Informatika Jurusan Teknik Informatika STMIK Subang, Jawa Barat, 2.
- Andriani, Anik.(2013). *Sistem Pendukung Keputusan Berbasis Decision Tree Dalam Pemberian Beasiswa Studi Kasus: Amik "Bsi Yogyakarta"*. Seminar Nasional : Program Studi Manajemen Informatika, AMIK BSI Jakarta
- Heriawati, Widodo Prabowo Pudjo, Handayanto Hahmadya Trias. (2013). *Penerapan Data Mining Dengan Matlab*. Bandung: Rekayasa Sains.
- Kania, Nia. (2006). *Stimulasi Tumbuh Kembang Anak Untuk Mencapai Tumbuh Kembang Yang Optimal*. Bandung seminar "Stimulasi Tumbuh Kembang Anak. Diakses dari [[http://pustaka.unpad.ac.id/wp-content/uploads/2010/02/stimulasi\\_tumbuh\\_kembang\\_anak\\_optimal.pdf](http://pustaka.unpad.ac.id/wp-content/uploads/2010/02/stimulasi_tumbuh_kembang_anak_optimal.pdf)](di akses tanggal 23 September 2014).
- Maulina, Inas Ulvy. (2014). *Klasifikasi Kanker Payudara Menggunakan Decision Tree Dengan Algoritma ID-3*. Jurnal teknik informatika dan ilmu komputer, vol.1, No.1
- Munawaroh, Holisatul. (2013). *Perbandingan Algoritma ID3 dan C5.0 dalam Identifikasi Penjurusan Siswa SM*. Jurnal Sarjana Teknik Informatika, Vol.1, No.1

Prasetyo, Eko. (2012). *Data Mining konsep dan aplikasi menggunakan matlab*. Yogyakarta: Andi.

Saputra, Rizal Amegia. (2014). *Komparasi Algoritma Klasifikasi Data Mining Untuk Memprediksi Penyakit Tuberculosis (Tb): Studi Kasus Puskesmas Karawang Sukabumi*. Seminar Nasional : Program Studi Manajemen Informatika, AMIK BSI Sukabumi.

Santosa, Budi. (2007). *Data Mining Terapan Dengan Matlab*. Yogyakarta: Graha Ilmu.

Shofiyatin, Lilik. (2011). *Penerapan Metode Decision Dengan Menggunakan ID3 Untuk Pembuatan Sistem Penilaian Kinerja Guru*. Skripsi. Malang : Fakultas Sains Dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim.

## **BIODATA PENULIS**

Nama : Mila Listiana

Nim : L200110047

Tempat Lahir : Ngawi

Tanggal Lahir : 7 Januari 1993

Jenis Kelamin : Perempuan

Agama : Islam

Pendidikan : S1

Jurusan : Informatika

Fakultas : Fakultas Komunikasi dan Informatika

Universitas : Universitas Muhammadiyah Surakarta

Alamat : RT.08 RW.06 Bulak Rejo, Katikan, Kedunggalar, Ngawi

Nomor Telepon : 085649157572

Email : milalistiana7@gmail.com